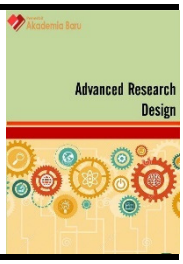
Penerbit Akademia Baru

Journal of Advanced Research Design

Journal homepage: www.akademiabaru.com/ard.html
ISSN: 2289-7984



Development of structure-activity modelling of carboxamides compounds for Aedes aegypti repellents

Open Access

Rosmahaida Jamaludin[1,*], Nurul Azmira Ibrahim[2], Hasmerya Maarof[3]

[1] UTM Razak School of Engineering and Advanced Technology, Universiti Teknologi Malaysia, Jalan Sultan Yahya Petra, 54100 Kuala Lumpur, Malaysia
[2] Kolej MARA Banting, Bukit Changgang, 42700 Banting, Selangor
[3] Department of Chemistry, Faculty of Science Universiti Teknologi Malaysia, 81310 Skudai Johor, Malaysia

ARTICLE INFO

Article history:
Received 21 May 2017
Received in revised form 17 July 2017
Accepted 2 August 2017
Available online 14 August 2017

ABSTRACT

Recently, research has been critically focused on finding new compounds with anti-repellent activity due to the rising of new types of mosquito-borne diseases. Mosquito repellents are the safer and cleaner alternative to fight the anthropods from bitten human skins, hence reduce the spread of diseases. This study investigated the relationships between biological activity and structure of carboxamides by using Quantitative Structure-Activity Relationship (QSAR) analysis. The data set used in this study comprised of 40 carboxamide compounds taken from the literature with their activities expressed as log PT (protection time). These compounds were split into training set for model building and test set for external validation using activity-based ranking method. The training set contained approximately 75% of the compounds while the remaining compounds were then used as the validation set to verify the accuracy of the model. DRAGON software was employed to generate molecular descriptors. The important relevant descriptors were further selected and reduced by using Genetic Algorithm (GA) as variable selection method. Two QSAR models were developed by combining GA method with two different modelling techniques that are multiple linear regressions (MLR) and partial least square (PLS). All the models are robust with good correlation coefficient (r2) greater than 0.6 and external validation r2test more than 0.5. Statistics of the GA-MLR model are r2 = 0.779 and r2test = 0.646. Whereas, the second model generated from GA and PLS shows good r2 with value of 0.775 and r2test = 0.563. These results could be useful in finding new, safe, and more effective repellents against Aedes Aegypti in a short time by providing guidance for further laboratory work as well as prediction of external compounds and help to understand the factors affecting their activity.

Keywords:
anti-repellent, mosquitoes, QSAR, DRAGON, GA, MLR, PLS

1. Introduction

Dengue is an infectious disease spread by the bite of female Aedes aegypti mosquito which is generally initiate from urban territories [1]. Urbanization [2] is a constructive aspect of Aedes aegypti

* *Corresponding author.*
E-mail address: Rosmahaida Jamaludin (rosma.kl@utm.my)

breeding and consequently aids the spread of the dengue virus known as DENV due to massive infrastructure expansion in the developed countries. According to Gubler [2], nearly 500 000 cases of dengue haemorrhagic fever (DHF) with 12 000 deaths has been recorded every single year.

Topical repellents are the most essential thing and play a vital role in disturbing the interaction between mosquitos and human that serves as a means of individual security. Revelation of new insect repellent active ingredients usually involves complex procedure with regularly changing new innovations. For instance, despite everything it takes around 10 years and by roughly 30 million to put up another new insect repellent for sale to the public. The utilization of Quantitative Structure–Activity Relationship (QSAR) ways to deal with repellent discovery is moderately exceptional [2]. In a standout amongst the latest studies, three dimensional (3D) QSAR was utilized effectively with CATALYST programming to create models for repellents based on pharmacophores [2]. Observation reveals that the study on prediction of repellents for the primary vector of dengue outbreak is infrequently carried out particularly in Malaysia. Therefore, it is vital to have a prediction model that could exceptional foresee those highly potential repellent for Aedes aegypti. Therefore, any innovation such as QSAR modelling that can enhance the effectiveness of the procedure is exceptionally significant to the business industry.

In this study, a data set of forty analogs of carboxamides taken from Suryanarayana, Pandey [3] work possessing anti-repellent activity towards Aedes aegypti against log PT was employed in the QSAR investigation to find its relationship with structures of carboxamides analogues. Each of carboxamides compounds is related to the ability to repel Aedes aegypti, the primary vector of dengue disease. This is one of the earlier QSAR work and several attempts by others [4-6] have been made to develop better models. According to Ma *et al*. [6], from the data set revisions of Suryanarayana *et al*. [3], the effectiveness of repellent activity is hypothesized to be from amide analogue but reported no numerical relationship. The amides represent the most commercially efficacious class of recent insect repellents. Bhattacharjee *et al*. [4] and Katritzky *et al*. [5] applied a three-dimensional chemical function based pharmacophore model for potent arthropod repellent activity using the same data set but variable selection in validation analysis was absence and over-fitting might occur as had been discussed by Kraker *et al*. [7].

The main focus of this project is the development and implementations of the molecular modelling techniques and statistical algorithms to process the biochemical data in drug discovery by constructing a few QSAR models using several methods. In order to develop better and robust predictive model, the built models were established by minimizing predictive error and over-training. These methods were applied to model the activity of carboxamide analogues which further validated by leave-one-out cross validation (LOOCV) for internal check and run external validations to test the predictive power of the built models. Hence, QSAR model is used to make predictions and explain the mechanism involved. This will be relevant to numerous industries such as biotechnological and pharmaceutical industries by providing them proven methods to cost and time effective in the development of new drugs as compared to conventional methods.

2. Methodology

2.1 Data Set

In the current report, diverse and high quality data sets are used for QSAR model building with known chemical and biological properties obtained from reliable literature. In this study, QSAR models were built to correlate protection time as anti-repellent activity towards Aedes Aegypti of carboxamide compounds and their molecular structures represented as molecular descriptors. Repellent activity is expressed in protection time (PT) and measured by hour. The data set used in

this work was taken from previous published report by Suryanarayana *et al*. [8] that comprised of forty benzamides, benzyl amides and cyclohexyl amide with their respective protection time. PT refers to the period of protection presented at given measured quantity of chemical repellent until two consecutive bites that are obtained at a 30-min interval. PT was reported as the average of several measurements. Interestingly, the optimal PT of carboxamides compounds were found to be between 0.08-6.00 hour at 30ºC. 40 compounds were under investigations and separated by two sets. Training set consists of 30 compounds and was used for model development, whereas another 10 compounds were selected to be in the test set by using activity-based ranking method due to their similarity in structural characteristics. The test set compounds were used to determine the predictive power of model generated from training set in order to plaid its validity.

2.2 Model Development

Initially, ChemDraw Ultra version 6.0 [9] was used to draw the 2-dimensional (2D) molecular structures of the compounds and converted into 3-dimensional (3D) by using Chem3D Ultra version 6.0 [9]. Subsequently, the energy of the generated three-dimensional structure of each carboxamide compound was minimized using semi empirical PM3 Hamiltonian in GaussView software package version 5.08 [10]. The next step was to calculate molecular descriptors that acts as independent variables for all compounds by using DRAGON software package version 5.4 [11] comprising of constitutional, topological, 2D autocorrelations and physical properties of molecules resulting in 3224 descriptors. The number of descriptors were further reduced to 918 by applying routine method of auto-scaling to zero mean and setting standard deviation to be less than 0.0001. In addition, objective feature selection was performed to exclude descriptors that were highly correlated and redundant. Such procedure includes removing constant and near-constant descriptors followed by pair-wise correlation coefficient that was greater than 0.90. The last elimination step was performing identical test by rejecting those having identical values of descriptors more than 90% of total compounds and as a result, 214 descriptors were left for further analysis.

The goal for feature selection is to identify the best subset of descriptors and also to decrease the descriptors pool to a sensible number that hold several redundant information by using Solo+Multivariate Image Analysis (MIA) software version 8.2.1 [12]. The subsequent method to obtain smaller number of descriptors that were information rich was subjective feature selection or more specifically GA. The parameter settings by GA were performed by replication runs of 30. In order to obtain an explicit final model, a stepwise approach was applied instead of directly using variables selected by GA [13]. The descriptors were selected based on the several highest frequencies of each variable in the top chromosome of each run. These frequencies were used as variables entered to generate robust model with good predictive power as displayed in figure 1. GA was embedded in regression methods such as MLR and PLS. Then, these combinations of variable selection and statistical methods were used to build several models.

QSAR equations also have been derived to provide a quantitative insight into correlation of the respective property and observed biological activity through regression coefficient. To obtain the best model, r2 was used to measure predictive performance, which corresponds to the degree of correlation existing between the predicted and experimental values. Consequently, important physico-chemical properties that accountable for an ideal insect repellent were identified. Finally, and the most important step was to validate the developed model internally and externally. Internal threefold cross-validation by leave-one-out (LOOCV) analysis was performed and all models were subjected to external validation using the test set compounds that are not contemplated during the model development. The QSAR models were internally validated by LOOCV analysis which repeats

the regression several times on data subsets by left out once at a time and r2cv value is calculated using the predicted values of the missing molecule. The resulting QSAR model can be utilized to select compounds that have similar structural attributes as the active compounds in the training set and they are expected to demonstrate repellent activity towards Aedes Aegypti. QSAR models have discovered extremely reasonable and practical applications by giving many important exploratory data in order to predict new compounds related to activity. In short, QSAR demonstrating is a valuable strategy for quickening improvement of medications, agro and fine chemicals, materials, and toxicology assessment.

3. Result and Discussion

3.1 Development of QSAR Model Using Log Protection Time (PT) as Biological Activity

Extensively, a data set of forty analogues of carboxamides possessing anti-repellent activity towards *Aedes aegypti* against log PT was employed in the QSAR investigation. Protection time is known as 'time to first bite' test and this labor-intensive method of repellent has been evaluated successfully to represent the efficacy of repellency [14]. As mentioned earlier, GA is applied to choose optimum quantity of descriptors in order to find a good model. Figure 1 displays the GA tools in SOLO+MIA software to identify significant descriptors to build robust models. As illustrated in the figure, GA frequency was used in selection of best response by entering the variables according to the frequency of selections.

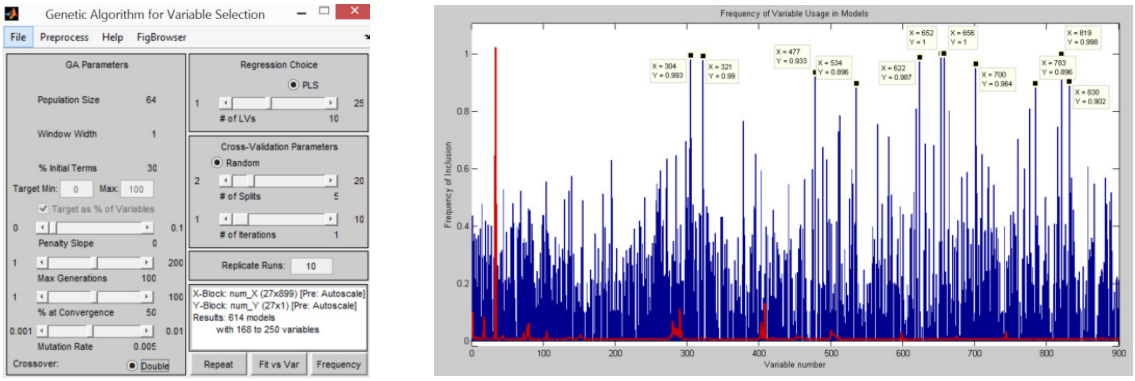


Fig. 1. Basic scheme of GA selection in SOLO+MIA v8.2.1

Therefore, by using the best descriptors selected by GA, two different QSAR models were constructed and represented as QSAR equations as shown in Equation 1 obtained by combining GA and MLR while Equation 2 by combining GA and PLS.

log PT = 2.66935(*MATS1e*)-36.8802(*JGI5*)-0.01584(*P_VSA_MR_5*)-
0.03385(*P_VSA_e_3*)+0.134746(*Eig11_EA(dm)*)-0.04907(*Eig12_EA(dm)*)-
0.29553(*CATS2D_06_DL*)+0.320977 (*CATS2D_07_DL*) +4.7466 (1)

log PT = 3.76927(*MATS1m*)+7.895667(*MATS1e*)-38.5313(*JGI5*)+0.144234(*SpMin4_Bh(s)*)-
0.01512(*P_VSA_MR_5*)-0.21671(*SM07_AEA(ri)*)+0.148946(*H_051*)-
0.4325(*CATS2D_06_DL*)+0.186317(*CATS2D_07_DL*)+3.48477 (2)

Descriptions for each descriptor selected to construct equations in all QSAR models is as summarized in Table 1. This led to a performance improvement as indicated by the statistical parameters presented in Table 2.

Table 1
Descriptions of selected variables

Descriptor	Group type	Description
MATS1e	2D autocorrelations	Moran autocorrelation of lag 1 weighted by Sanderson electronegativity
MATS1m	2D autocorrelations	Moran autocorrelation of lag 1 weighted by mass
JGI5	2D autocorrelations	mean topological charge index of order 5
P_VSA_MR_5	P_VSA-like descriptors	P_VSA-like on Molar Refractivity, bin 5
P_VSA_e_3	P_VSA-like descriptors	P_VSA-like on Sanderson electronegativity, bin 3
Eig11_EA(dm)	Edge adjacency indices	eigenvalue n. 11 from edge adjacency mat. weighted by dipole moment
Eig12_EA(dm)	Edge adjacency indices	eigenvalue n. 12 from edge adjacency mat. weighted by dipole moment
CATS2D_06_DL	CATS 2D	CATS2D Donor-Lipophilic at lag 06
CATS2D_07_DL	CATS 2D	CATS2D Donor-Lipophilic at lag 07
SpMin4_Bh(s)	Burden eigenvalues	smallest eigenvalue n. 4 of Burden matrix weighted by I-state
SM07_AEA(ri)	Edge adjacency indices	spectral moment of order 7 from augmented edge adjacency mat. weighted by resonance integral

Table 2
Summary of predictive performance of two different methods

Model	Method	RMSEC	RMSECV	RMSEP	r^2	r^2_{cv}	r^2_{test}
1	GA-MLR	0.231	0.356	0.258	0.779	0.527	0.646
2	GA-PLS	0.233	0.386	0.309	0.775	0.482	0.564

The best model was using hybrid approach of GA and MLR, possessing correlation coefficients r^2_{test} and *RMSEP* of 0.646 and 0.258 respectively, as obtained from external validation test. The high value of r^2_{test} and low value of *RMSEP* specified a more robust model.

3.4 Interpretation of Descriptors

The top two 2D descriptors obtained through feature selection were shown to be useful in the construction of robust QSAR models and these descriptors were greatly shown its ability to explain all of the variations with structures of carboxamides that possess anti-repellent activity. It is interesting to note that the descriptor *CATS2D_07_DL* contributed significantly in both QSAR models. This descriptor contributed the largest coefficient value followed by *Eig11_EA(dm)*. The details on the selected variables are presented in Table 1. Dipole moment is known as one of electronic property that have an important role in affecting activity of repellent. It is related to the polarity of the molecule when atoms in the molecule share electrons unequally and resulted in permanent dipole moment.

According to Ma *et al*. [6], the higher the magnitude of dipole moment of a compound, the more likely it is to be a very good indicator to hydrophobicity. From their findings, the authors also concluded that lipophilicity or hydrophobicity for this class of compounds is essential and could be

the important factors to be an active repellent. Figure 2 represent one of the carboxamide compounds from data set and hydrophobicity area is as labelled. Study of Skinner and Johnson [15] shows a relationship with necessity of lipid solubility for insects' olfactory sensation and hydrophobicity of repellent as significant aspect too.

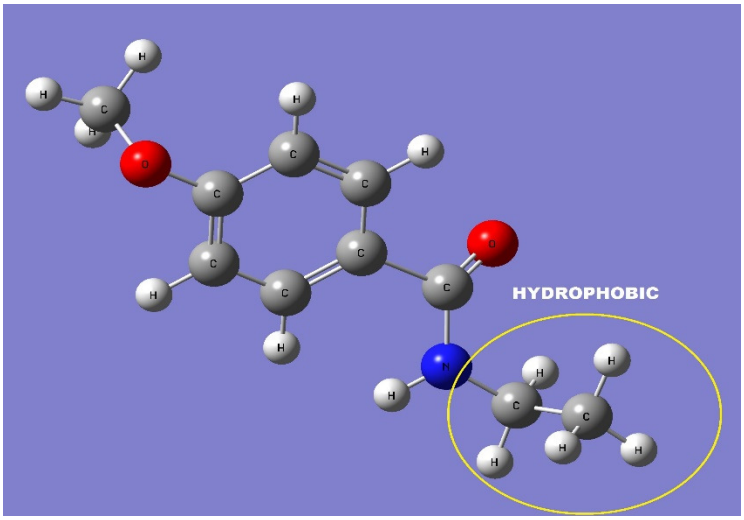


Fig. 2. Gaussian 3D view of compound 1 from data set

The combination of GA and MLR to find the most stable model is more useful since it can greatly improves the predictive ability of the model and significantly reduced the number of descriptors. The selected descriptors are known to be the most informative variables that correspond to the structures. This model gives good statistical parameters and has the optimal number of components (*N*= 8) with r^2_{test} of 0.646, which is better than the other value from the GA-PLS model. Thus, the derived model from GA-MLR method as represented in Table 2 can be used in prediction of PT for new compounds, in the homologous series. The regression model consists of regression coefficient reported in brackets. *Log* PT predicted by GA-MLR model for ten compounds in the test set was consistent with the measured data and in perfect agreement with each other as evident in Table 3.

Table 3
Measured and predicted activities of anti-repellent activities for test set compounds

Compound	Log PT		
	Measured	Predicted	Residual
9	1.4781	1.8587	-0.3806
10	1.7796	1.3129	0.4667
13	2.3794	2.6412	-0.2618
16	1.6074	1.6521	-0.0447
20	1.9308	1.9775	-0.0467
23	2.2566	2.4094	-0.1528
27	2.2311	2.1492	0.0819
32	2.1163	2.5431	-0.4268
33	2.558	2.5051	0.0529
34	1.7801	1.8761	-0.0960

4. Conclusion

The goal of this work was to investigate a strategy aimed at producing a robust model that can predict the ability of mosquito repellents from carboxamides analogs. The repellency class data of a

set of carboxamides from the Suryanarayana *et al*., [8] were used to develop suitable QSAR models to predict new repellent structures. In this present work, predictive models for repellency were built using QSAR methods in combination with regression tools and GA. Analysis of the 2D properties of carboxamide structures revealed two parts of the molecule that were significant to repellent activity and probably involved in insect receptor-repellent interactions. Dipole moment and lipophilicity were two most informative descriptors selected by GA and well described the chemical reactivity of *log* protection time with repellency. In summary, the combination approach with GA-MLR analysis to build best model of prediction for *log* PT bioactivity is reliable and robust which may be able to predict molecular and electronic properties of chemicals from amides homologous series that result in anti-repellent towards *Aedes aegypti*.

References

[1] Smith, C. E. G. "The history of dengue in tropical Asia and its probable relationship to the mosquito Aedes aegypti." *Journal of Tropical Medicine and Hygiene* 59, no. 10 (1956): 243-51.
[2] Gubler, Duane J. "Epidemic dengue/dengue hemorrhagic fever as a public health, social and economic problem in the 21st century." *Trends in microbiology* 10, no. 2 (2002): 100-103.
[3] Suryanarayana, M. V. S., K. S. Pandey, Shri Prakash, C. D. Raghuveeran, R. S. Dangi, R. V. Swamy, and K. M. Rao. "Structure-activity relationship studies with mosquito repellent amides." *Journal of pharmaceutical sciences* 80, no. 11 (1991): 1055-1057.
[4] Bhattacharjee, Apurba K., Watanaporn Dheranetra, Daniel A. Nichols, and Raj K. Gupta. "3D pharmacophore model for insect repellent activity and discovery of new repellent candidates." *Molecular Informatics* 24, no. 5 (2005): 593-602.
[5] Katritzky, Alan R., Dimitar A. Dobchev, Indrek Tulp, Mati Karelson, and David A. Carlson. "QSAR study of mosquito repellents using Codessa Pro." *Bioorganic & medicinal chemistry letters* 16, no. 8 (2006): 2306-2311.
[6] Ma, D., APURBA K. Bhattacharjee, R. K. Gupta, and JEAN M. Karle. "Predicting mosquito repellent potency of N, N-diethyl-m-toluamide (DEET) analogs from molecular electronic properties." *The American journal of tropical medicine and hygiene* 60, no. 1 (1999): 1-6.
[7] Kraker, Jessica J., Douglas M. Hawkins, Subhash C. Basak, Ramanathan Natarajan, and Denise Mills. "Quantitative Structure–Activity Relationship (QSAR) modeling of juvenile hormone activity: Comparison of validation procedures." *Chemometrics and intelligent laboratory systems* 87, no. 1 (2007): 33-42.
[8] Suryanarayana, M. V. S., K. S. Pandey, Shri Prakash, C. D. Raghuveeran, R. S. Dangi, R. V. Swamy, and K. M. Rao. "Structure-activity relationship studies with mosquito repellent amides." *Journal of pharmaceutical sciences* 80, no. 11 (1991): 1055-1057.
[9] Ultra, ChemDraw. "6.0, CambridgeSoft Corporation, 100 CambridgePark." *Cambridge, MA* 2140.
[10] Dennington, Roy, Todd Keith, and John Millam. "Semichem Inc." *Shawnee Mission KS, GaussView, Version* 5 (2009).
[11] Mauri, Andrea, Viviana Consonni, Manuela Pavan, and Roberto Todeschini. "Dragon software: An easy approach to molecular descriptor calculations." *Match* 56, no. 2 (2006): 237-248.
[12] Toolbox, P. L. S. "5.8, Eigenvector Research." *Inc., USA* (2010).
[13] Leardi, Riccardo, and Amparo Lupianez Gonzalez. "Genetic algorithms applied to feature selection in PLS regression: how and when to use them." *Chemometrics and intelligent laboratory systems* 41, no. 2 (1998): 195-207.
[14] Gupta, Raj K., and Apurba K. Bhattacharjee. "Discovery and design of new arthropod/insect repellents by computer-aided molecular modeling." *Insect Repellents: Principles, Methods, and Uses* (2007): 195-228.
[15] Skinner, W. A., and H. L. Johnson. "The design of insect repellents." *Drug design* 10 (1980): 277-305.